EPiC Series in Computing

Volume 107, 2025, Pages 272–281

Proceedings of EUNIS 2025 annual congress in Belfast

Preserving Research Grant Data: Building an OAIS-Compliant Open Archive

Abdelaziz Bouras[*,1], Hira Naseem[1] and Abdelhak Belhi[2,1]

[1]Qatar University, Doha, Qatar

[2]Joaan Bin Jassim Academy for Defence Studies, Al Khor, Qatar

abdelaziz.bouras@qu.edu.qa, hn2000131@qu.edu.qa, abdelhak.belhi@jbj.edu.qa

Abstract

The preservation of research grants' data is essential for long-term accessibility, auditing, and informed decision-making. This paper presents the implementation of the Open Archival Information System (OAIS) model for archiving research grants' data. The proposed system ensures structured storage, retrieval, and management of critical research outputs, including reports, publications, and financial records. While the implementation has addressed key challenges such as data ingestion, metadata management, and security, certain aspects require further refinement to optimize scalability and accessibility. Despite these challenges, the OAIS-based approach provides a robust foundation for future-proofing research data, ensuring its availability for strategic planning, performance evaluation, and institutional decision-making. The study highlights lessons learned from the initial deployment and outlines enhancements needed to improve system efficiency. Ultimately, this model serves as a sustainable framework for research data archiving, contributing to a more data-driven research ecosystem.

1 Introduction

In the digital age, the rapid expansion of data has necessitated robust strategies for its preservation, accessibility, and security (Cliggett, 2013). Governments, academic institutions, and industries generate vast amounts of digital assets, ranging from research findings and scientific data to administrative records and intellectual property. Without a structured approach to archiving, valuable information risks being lost due to technological obsolescence, hardware failures, or inadequate storage practices (Adu et al., 2016). This highlights the critical need for digitization and the development of sustainable archival frameworks to ensure long-term data integrity and retrieval.

* Corresponding Author

Digitization has transformed the way organizations manage information. While digitization enhances accessibility and efficiency, it also introduces challenges related to data authenticity, preservation, and security. Digital archives are susceptible to corruption, unauthorized modifications, and loss due to software dependencies or changing file formats. To mitigate these risks, standardized archival models are essential for ensuring the long-term usability and trustworthiness of stored records (Adu et al., 2016). Additionally, archival initiatives in various domains, including engineering, have explored methods for preserving information within specific contexts (Bahloul et al., 2008).

Research grants' data accumulates across multiple phases, making it essential to implement a structured system to store, retrieve, and analyze this information systematically. To manage this complex process and vast amount of data, we have developed the Research Tracking System (RTS) (Bouras et al., 2022). This system is evolving ever since which now (Bouras et al., 2025), it facilitates reviewer feedback, and proper evaluation of research outcomes. But it does not provide a structured mechanism for long-term data preservation and historical analysis across the full research lifecycle. Additionally, the Pre-Award Team currently uses external tools to receive, process and store its data. Due to retention policies requiring data deletion every three years, a reliable archival mechanism is necessary to preserve essential records beyond this period. Thus, a comprehensive archiving system is required to address these gaps. Therefore, we propose an archival system based on the OAIS reference model (Lavoie, 2014), a widely adopted framework for long-term digital preservation.

Archiving research data poses several challenges, such as standardizing metadata, ensuring long-term data integrity, managing large, diverse datasets, and enforcing secure access. Without a clear framework, institutions risk data loss, fragmentation, or inaccessibility. A robust archiving system enables effective grant management, strategic evaluation, and long-term optimization, while also supporting data-driven decisions and reinforcing a strong research environment.

Qatar University's Research Support Department is developing an OAIS-compliant archiving system to improve research grant data management and accessibility. Effective archiving supports auditing, strategic planning, and long-term preservation. This case study explores the system's architecture, metadata handling, security, and preservation standards, offering best practices for digital archiving in academia.

The paper is structured as follows: Section 1 outlines the research problem and objectives; Section 2 provides background and related work; Section 3 presents our OAIS-based modeling approach; Section 4 discusses results; and Section 5 concludes with key insights and future directions.

2 Background

2.1 Overview of digital reservation and archival systems

Digital preservation is a critical component of research data management, ensuring that valuable scientific and administrative records remain accessible, secure, and verifiable over time. Archival systems provide structured methodologies for storing, retrieving, and maintaining digital assets, addressing risks such as data corruption, technological obsolescence, and unauthorized access.

Without a systematic archival framework, research institutions face challenges in data integrity, metadata standardization, and interoperability. The emergence of internationally recognized frameworks, such as the OAIS, has provided a structured model for addressing these challenges by defining essential functions, such as data ingestion, preservation planning, and dissemination.

2.2 OAIS reference model and its adoption in research data management

The OAIS reference model, developed by the Consultative Committee for Space Data Systems (CCSDS) and standardized as ISO 14721, provides a comprehensive approach to long-term digital preservation. OAIS consists of six key functional components: INGEST, Archival Storage, Data Management, Administration, ACCESS, and Preservation Planning. The Ingest function handles data submission and metadata generation, while Archival Storage ensures secure and long-term preservation. Data Management organizes Metadata for efficient retrieval, and Access enables users to retrieve archived content. Preservation Planning ensures data remains usable despite technological changes, while Administration oversees overall system operations. These components ensure the long-term preservation and accessibility of digital information. The conceptual model is presented in Figure 1. Another core aspect of OAIS is its structured handling of information packages with three actors PRODUCERS, MANAGEMENT and CONSUMERS. Data enters the system as a Submission Information Package (SIP), which includes raw data and metadata provided by the PRODUCER. Once ingested, it is transformed into an Archival Information Package (AIP), by MANAGEMENT. The AIP contains both the original data and preservation metadata for long-term management. When CONSUMER requests archived data, the system generates a Dissemination Information Package (DIP), delivering the requested content in a user-accessible format. These standardized processes ensure that archived research data can be securely maintained and retrieved in a meaningful way, even as technology evolves.

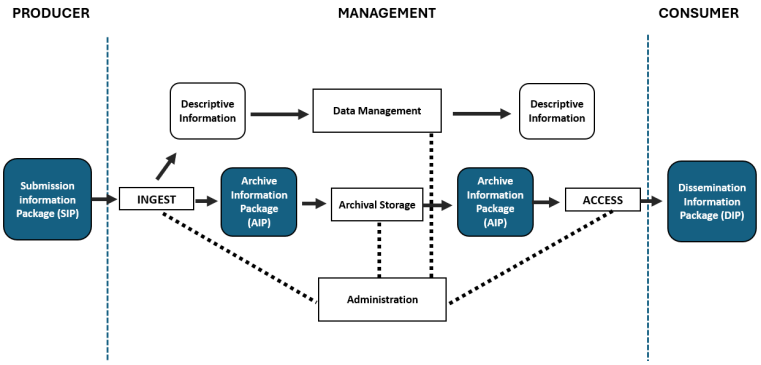

Figure 1: OAIS Reference Model. The dashed partition lines separate the functional responsibilities of each actor, while the arrows represent the flow of data across components. The dotted connections indicate administrative oversight across functional areas

Several institutions have integrated OAIS principles into their digital preservation workflows. Its adoption has facilitated the standardization of archival policies, improved data discovery mechanisms, and strengthened security frameworks through fixity checks and access control mechanisms.

2.3 Challenges in archiving scientific and research grants' data

Despite the benefits of archival systems, research institutions face multiple challenges in preserving grant-related data. These include handling diverse data formats, managing growing data volumes, ensuring security and integrity, and addressing long-term accessibility amid technological obsolescence. Effective preservation requires strategies like format migration and emulation. Many studies have adopted the OAIS model, adapting it to meet these specific institutional needs such as (Caplan, 2010; Farquhar & Hockx-Yu, 2008; Tansley et al., 2003) and so on.

The National Space Science Data Center (NSSDC) used the OAIS Reference Model to structure its migration of legacy data sets from old magnetic tape onto digital linear tape (DLT) (Ball, 2006). Dark

Archive in the Sunshine State (DAITSS), includes bit-level preservation, format normalization, and forward format migration, ensuring long-term data integrity. A second version, DAITSS 2, was planned as RESTful web services, making the system openly accessible to other institutions (Caplan, 2010).

The European Planets Project focused on addressing technological obsolescence to ensure long-term access to digital assets (Farquhar & Hockx-Yu, 2008). The UK Data Archive, a major repository for social and economic research, emphasizes structured metadata and compliance with legal and funding requirements (Loesch, 2015). CERN applied OAIS principles to maintain data authenticity, enable fixity checks, and support scalable, secure dissemination(Meur & Tarocco, 2019).

The MathArc project developed tools to enable sharing and storage of digital objects across OAIS-based repositories, regardless of system differences. Other repositories and academic institutions have also adopted the OAIS model for preserving theses, research outputs, and funding records, highlighting its value for institutional archiving (Ball, 2006; Qasim et al., 2018)

Although Bleakly (2002) predates recent implementations, it emphasizes a lasting principle: the OAIS model must be adapted to fit institutional contexts. As seen in Section 2.3, each implementation customized OAIS to meet specific organizational and technical needs. Key insights include automating ingest pipelines, capturing usable metadata, ensuring system interoperability, performing regular fixity checks, and designing user-friendly access interfaces. These findings highlight OAIS's flexibility and the ongoing need for refinement in archival systems.

3 Use Case

Administrative offices support research by providing essential services, as shown in Figure 2. Research funding programs generate extensive data, including proposals, budgets, approvals, contracts, reports, publications, patents, and student records. These datasets are vital for auditing, evaluation, and strategic planning. However, without a structured archival system, preserving and managing this information remains a challenge.

Figure 2: Services provided during project lifecycle. Data of dotted box has been archived in this study

At QU research grant project follows a structured three-phase lifecycle, each generating valuable data as depicted in Figure 3:

1. *Pre-Award Phase*: This phase initiates with the call for proposals, during which Lead Project Investigators (LPIs) submit their research proposals. These proposals detail the preliminary research idea, expected outcomes, research commitments, potential future impact, and the required budget.

2. *Contract & Compliance Phase*: Once projects receive funding approval, the Contract and Compliance Team is responsible for ensuring that all legal, financial, and administrative requirements are met.
3. *Post-Award Phase*: LPIs must submit a progress and final reports detailing their research advancements and any challenges encountered. A review committee (Post-Award Scientific Committee), composed of field experts, evaluates these reports, offering insights, recommendations.

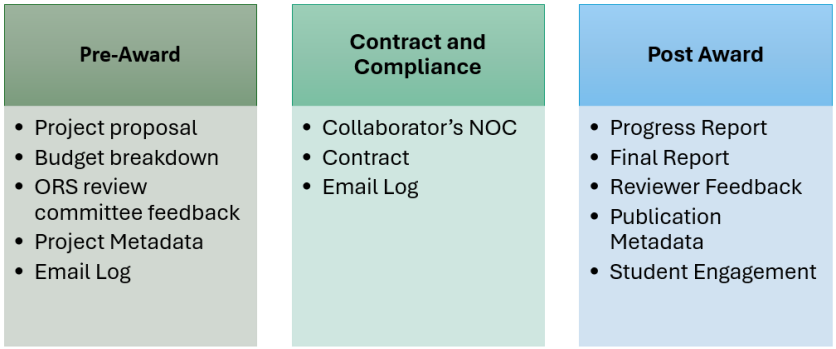

Figure 3: Data generated at different phases of projects

To address this, we are implementing an OAIS-compliant archiving system that will systematically INGEST, store, manage, and provide access to this research data. According to OAIS, A data object is created from either a physical object or a digital object. This data object is then supplemented with a knowledge base, specific to the designated community and representation information (such as metadata or documentation). This combination transforms the data object into an information object. Subsequently, the information object is further enhanced with Content Information, Packaging Information, Descriptive Information, and Preservation Description Information. These modifications result in the formation of an information package, ready for storage and retrieval within the archive. This whole process is depicted in Figure 4.

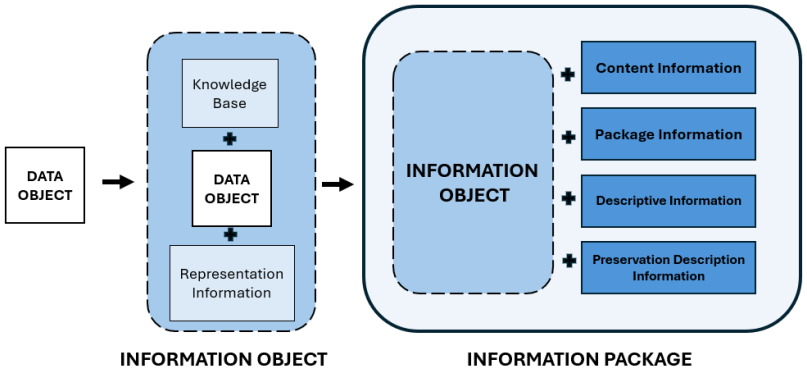

Figure 4: Generalized process of converting a Data Object into an Information Package. The term 'Information Package' is used here in a generic sense to represent any of the OAIS-defined information packages

In our archiving system, when a SIP is uploaded, it undergoes a validation process to confirm that it meets the required format for the phase in which it is submitted. The system ensures that each SIP adheres to the expected structure before further processing. If the SIP fails validation, an error is prompted, preventing further actions until the issue is resolved. Once the SIP is validated, a

cryptographic hash is generated and appended, transforming it into an AIP. This AIP retains the original contents of the SIP, with added metadata, including the hash for integrity verification. The finalized AIP is securely stored on the server. The system follows a structured process to handle data, converting it into a SIP and verifying its integrity before committing it as an AIP.

- **Query the Data**: Initially, data is extracted from the database using SQL queries.
- **Transform the Data**: The tabular data is then projected to data cleaning and pre-processing and data verification using Python Script. Then it is transformed into a JSON object using built-in tools.
- **Hash the Data**: Next, separate hashes are computed for both the tabular data and the relevant PDF files, which are then appended to the JSON Objects. This is implemented using the Laravel framework.
- **Package the Data**: Additional metadata, including content information and preservation description information etc., as mentioned in Figure 4, is also incorporated.
- **Verify the Data**: A verified and authorized user then requests to store the data in the archival system. At this stage, the system verifies the hash by recalculating the hash. If it matches the appended hash, the data is considered safe. Otherwise, the system detects a data integrity failure.
- **Upload the Data**: Once verified, the data is transformed back into its tabular format and stored as an AIP in the database within the archival system, along with its metadata and other relevant information.

We are currently implementing and testing our DIP strategies. The data retrieval process begins with user authentication and access control. To enhance data integrity, blockchain integration is being explored by storing AIP hashes on a decentralized ledger to prevent tampering. Initial tests are successful, with full integration planned. The system uses Fiber SDK—a lightweight, modular framework for interacting with Hyperledger Fabric via high-level APIs. Figure 5 illustrates this interaction.

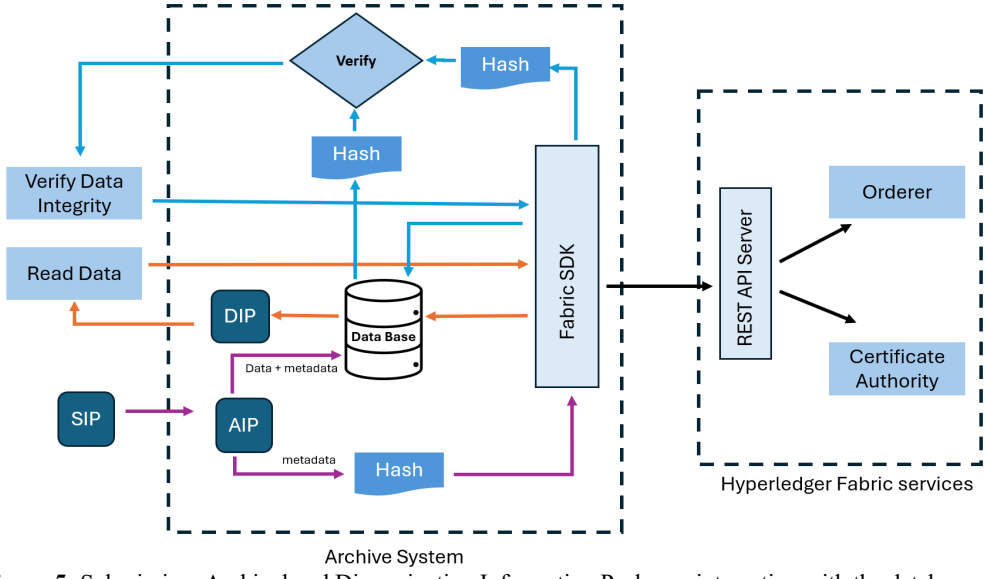

Figure 5: Submission, Archival and Dissemination Information Packages interacting with the database and Fabric SDK. Cryptographic hashes of AIPs are stored on Hyperledger Fabric, ensuring data integrity. User access is controlled through a Certificate Authority, and real-time hash verification against stored blockchain hashes guarantees tamper-proof data retrieval

3.1 OAIS-compliant system development

The process was initiated by engaging key stakeholders, Pre-Award, Post-Award, and Contract & Compliance Teams, to understand workflows and archival needs. Objectives included identifying data types for archiving, defining metadata standards, setting retention policies, and addressing security concerns. This aligns with the OAIS INGEST function, establishing how data is collected, structured, and prepared for long-term storage.

In the system design phase, we defined metadata standards, structured storage options (centralized, cloud, or hybrid), and created a unique Project ID system to link research records. User access levels were also established to meet governance requirements, aligning with OAIS Data Management and Archival Storage functions.

During development, we implemented key components: structured data ingestion, automated metadata assignment, and validation mechanisms. Integration with tools like RTS supports seamless data migration. This phase reflects OAIS Ingest and Data Management functions. A private blockchain model is used to track and verify archival processes, ensuring security and control.

4 Discussion

This section presents the preliminary outcomes from the partial implementation of the OAIS-based archival system for research grants' data at QU. As the system is still in its development stages, the focus here is on the steps taken, the current integration of key OAIS components, and the challenges encountered.

The implementation of the OAIS-based archival system has progressed through several key stages. First, the Ingest function was established, with research grants' data for the Pre-Award phase being successfully ingested into the system. Data types include research proposals, budget breakdowns, progress reports, final reports, and publications. Metadata for these documents was generated and stored, with data validation mechanisms applied to ensure consistency and accuracy. Currently, 100+ reports have been ingested, and the system is prepared to handle future data inflows.

4.1 Blockchain integration for data integrity

In selecting a private blockchain architecture (Hyperledger Fabric with RAFT consensus), our intent was not to rely on blockchain's often overstated claim of immutability alone, but to leverage its ability to provide a transparent, tamper-evident log of interactions.

As part of the Fixity and Security Enhancements, a partial integration of blockchain was implemented to ensure the integrity and immutability of research grants' data. The blockchain framework, utilizing Hyperledger Fabric, was chosen to optimize transaction throughput while maintaining the security of metadata.

4.2 Challenges encountered

While the benefits of blockchain integration are clear, the process presents several practical challenges. One of the primary concerns is scalability, as blockchain systems require substantial storage and computational power to maintain and verify transactions. Integrating blockchain with our existing RTS posed additional challenges. Another challenge is transaction throughput and latency. Blockchain transactions require validation and recording across multiple nodes, which can lead to delays, especially during peak periods.

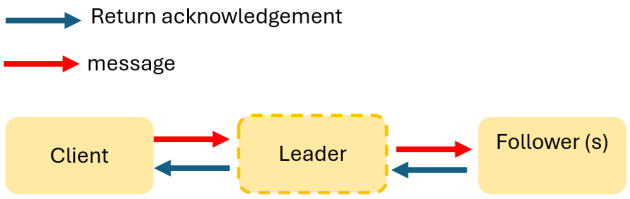

Figure 6: Raft Consensus Block Diagram

Finally, there are privacy and legal concerns related to blockchain's immutability. Storing sensitive research data directly on a blockchain could expose confidential information, so we only store metadata (such as file hashes) on the blockchain, while the actual content is encrypted and stored securely. This ensures that even if the blockchain is accessed, the data remains protected.

- **Scalability and Storage**: The blockchain's inherent storage constraints were highlighted during the initial integration. While storing only metadata on the blockchain has alleviated some of these concerns, issues related to the volume of data generated by research grants persist.
- **Data Privacy and Compliance**: Ensuring compliance with legal frameworks, such as GDPR, has been an ongoing challenge. The system's hybrid approach of storing only metadata on the blockchain, while the actual data resides off-chain in encrypted storage, has been crucial in addressing this issue. Implementation of smart contracts is in pipeline to ensure proper access controls and data retention policies.
- **System Integration:** These challenges are being mitigated through the development of API-based integrations and a middleware layer to streamline data flow between RTS and the blockchain.
- **Time Complexity:** o address this, we implemented a permissioned blockchain using Hyperledger Fabric with the Raft consensus protocol. Raft, suited for small networks, enables efficient transaction processing through a leader-based consensus mechanism. The leader sends log entries to follower nodes, which confirm receipt to maintain consistency. This approach reduces communication overhead compared to PBFT. Raft's time complexity varies by phase. Leader election uses a randomized timeout mechanism, with average complexity O(T), where T is the election timeout. Log replication takes O(1) per entry and O(n) overall, where n is the number of follower nodes. Thus, Raft's complexity ranges from O(1) in the best case to O(nT) in the worst case. A conceptual view is presented in Figure 6**.** RAFT has less communication overhead compared to Practical Byzantine Fault Tolerance (PBFT), as presented in Table 1.

Consensus Algorithm	Leader Election Complexity	Log Replication Complexity	Fault Tolerance	Communication Overhead
Raft	O(nT)		Crash Fault Tolerance (CFT)	Low
Paxos	O(n2)	O(n2)	CFT	High
PBFT	O(n2) (Alqahtani, n.d.)	O(n2)	Byzantine Fault Tolerance (BFT)	Very High

Table 1: Complexity of Consensus algorithm

4.3 Early-stage insights

While the system is still in its early stages, several important insights have emerged from the partial implementation. The blockchain-based solution has proven effective at preserving the integrity of research grants' data, with early tests demonstrating that metadata stored on the blockchain remains tamper-proof. This ensures that once research grants' data is ingested into the system, it is securely archived with a verifiable history. Initial user feedback from administrators indicates that the interface is intuitive, with an emphasis on searchability and metadata retrieval.

5 Conclusion

This study presented the development of an archiving system for research grants' data, designed using the OAIS reference model as a guiding framework. The system addresses the need for structured, long-term preservation and accessibility of research data generated throughout the Pre-Award, Award & Compliance, and Post-Award phases. By aligning with OAIS principles, the system ensures that data is systematically ingested, securely stored, efficiently managed, and readily accessible to stakeholders, including researchers, administrators, and funding bodies.

One of the key contributions of this system is the integration of blockchain-based security mechanisms to enhance data integrity and authenticity. By leveraging Hyperledger Fabric, the system records metadata hashes on an immutable ledger, ensuring that research grant records remain tamper-proof while complying with privacy regulations such as GDPR. Additionally, the system incorporates fixity checks, encryption, and metadata indexing, reinforcing its reliability as a trusted archival solution for funded research projects.

Acknowledgement

The Authors thank Qatar University Research Grants & Contracts office for its support through QUEX-ORS-RTS-24/25-1 project. The statements made herein are the responsibility of the authors.

References

Adu, K. K., Dube, L., & Adjei, E. (2016). Digital preservation: the conduit through which open data, electronic government and the right to information are implemented. *Library Hi Tech*, *34*(4), 733–747.

Alqahtani, S. (n.d.). *Bottlenecks in Blockchain Consensus Protocols*. *3*.

Bahloul, K., Buzon, L., & Bouras, A. (2008). *Archival Initiatives in the Engineering Context BT - Global Design to Gain a Competitive Edge* (X.-T. Yan, W. J. Ion, & B. Eynard (eds.); pp. 313–321). Springer London.

Ball, A. (2006). *Briefing Paper: the OAIS Reference Model*. *kim12rep001ab10*, 1–18. http://www.ukoln.ac.uk/projects/grand-challenge/papers/oaisBriefing.pdf

Bouras, A., Al-Maadeed, M., Naseem, H., Hussain, S., Agouni, A., & Al-Salem, M. (2022). Development of a Research Tracking System for Higher Education Institution Research Grants. *EPiC Series in Computing*, *86*, 109–120. https://doi.org/10.29007/jxv4

Bouras, A., Naseem, H., Agouni, A., Al-meer, S., & Al-maadeed, M. (2025). *Research Management : Evolving Reporting and Scoring Capabilities in Research Tracking Systems for Higher Education Institutions 1 Introduction*. *105*, 49–57.

Caplan, P. (2010). DAITSS, an OAIS-based preservation repository. *Proceedings of the 2010 Roadmap for Digital Preservation Interoperability Framework Workshop*. https://doi.org/10.1145/2039274.2039291

Cliggett, L. (2013). Qualitative data archiving in the digital age: Strategies for data preservation and sharing. *The Qualitative Report*, *18*(24), 1.

Farquhar, A., & Hockx-Yu, H. (2008). Planets: Integrated services for digital preservation. *Serials*, *21*(2), 140–145. https://doi.org/10.1629/21140

Lavoie, B. (2014). Information System (OAIS) Reference Model : Introductory Guide. *Technology Watch Reports*, 37.

Loesch, M. (2015). UK Data Archive. *Technical Services Quarterly*, *32*. https://doi.org/10.1080/07317131.2015.1000736

Meur, J.-Y., & Tarocco, N. (2019). The obsolescence of Information and Information Systems CERN Digital Memory project. *EPJ Web of Conferences*, *214*, 9003. https://doi.org/10.1051/epjconf/201921409003

Qasim, U., Davis, C., Garnett, A., Marks, S., & Mosseburger, M. (2018). *Research data preservation in Canada: a white paper*. *April*, 1–16. https://portagenetwork.ca/wp-content/uploads/2018/05/Portage-PEG-WhitePaper-EN.pdf

Tansley, R., Bass, M., Stuve, D., Branschofsky, M., Chudnov, D., McClellan, G., & Smith, M. (2003). *The DSpace Institutional Digital Repository System: Current Functionality* (Vol. 2003). https://doi.org/10.1109/JCDL.2003.1204846

Author biographies

Prof. Abdelaziz Bouras is the Pre-Award Manager at the Research Support Department of Qatar University and Professor at the Computer Science and Engineering Department of the College of Engineering (CENG). He is leading the WG5.1 group at the International Federation of Information Processing (IFIP). He contributed to the publication of several books, such as the recent one titled "*The Sustainable University of the Future*" which discusses the rapid changes taking place within institutions of higher education. Prof. Bouras is currently coordinating the deployment of research information systems at the Research Support Department of Qatar University.

Hira Naseem completed her M.Sc. in Computing from Qatar University and her B.Sc. in Computer Science from the University of Peshawar, Pakistan. Her current interests include Machine Learning and AI-based solutions. She was recently awarded a Graduate Assistant position. She currently works as Research Assistant at the Research Support Department at Qatar University, where she contributes to developing frameworks and AI-based services for research tracking, smart archiving, etc.

Dr. Abdelhak Belhi, PhD, is an assistant professor at Joaan Bin Jassim Academy for Defence Studies. He recently served as adjunct research assistant professor at Qatar University. He received his Ph.D. in computer science and machine learning from the University of Lyon, France. Dr. Belhi has published numerous research articles in high-ranking peer-reviewed journals. Dr. Belhi co-edited a book on data analytics in the digital cultural domain and has served as a reviewer for leading journals in machine learning and deep learning. His current research interests include artificial intelligence, machine learning, cybersecurity, and blockchain.